DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs

Craig Knox[1], Vivian Law[2], Timothy Jewison[1], Philip Liu[3], Son Ly[2], Alex Frolkis[1], Allison Pon[1], Kelly Banco[2], Christine Mak[2], Vanessa Neveu[1], Yannick Djoumbou[3], Roman Eisner[1], An Chi Guo[1] and David S. Wishart[1,2,3,4,*]

[1]Department of Computing Science, University of Alberta, Edmonton, AB, Canada T6G 2E8,
[2]Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada T6G 2N8,
[3]Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada T6G 2E8 and
[4]National Institute for Nanotechnology, 11421 Saskatchewan Drive, Edmonton, AB, Canada T6G 2M9

ABSTRACT

DrugBank (http://www.drugbank.ca) is a richly annotated database of drug and drug target information. It contains extensive data on the nomenclature, ontology, chemistry, structure, function, action, pharmacology, pharmacokinetics, metabolism and pharmaceutical properties of both small molecule and large molecule (biotech) drugs. It also contains comprehensive information on the target diseases, proteins, genes and organisms on which these drugs act. First released in 2006, DrugBank has become widely used by pharmacists, medicinal chemists, pharmaceutical researchers, clinicians, educators and the general public. Since its last update in 2008, DrugBank has been greatly expanded through the addition of new drugs, new targets and the inclusion of more than 40 new data fields per drug entry (a 40% increase in data 'depth'). These data field additions include illustrated drug-action pathways, drug transporter data, drug metabolite data, pharmacogenomic data, adverse drug response data, ADMET data, pharmacokinetic data, computed property data and chemical classification data. DrugBank 3.0 also offers expanded database links, improved search tools for drug–drug and food–drug interaction, new resources for querying and viewing drug pathways and hundreds of new drug entries with detailed patent, pricing and manufacturer data. These additions have been complemented by enhancements to the quality and quantity of existing data, particularly with regard to drug target, drug description and drug action data. DrugBank 3.0 represents the result of 2 years of manual annotation work aimed at making the database much more useful for a wide range of 'omics' (i.e. pharmacogenomic, pharmacoproteomic, pharmacometabolomic and even pharmacoeconomic) applications.

INTRODUCTION

Historically most of the known information on drugs, drug targets and drug action has resided in books, journals and expensive commercial databases. Over the past 5 years this situation has changed quite dramatically. Now most drug and drug target data is freely available over the internet. The first on-line database to break the commercial 'stranglehold' on drug information was the Therapeutic Target Database (TTD), which was released in 2002 and then updated in 2010 (1). Over the years, other drug-specific databases have emerged, including PDTD (2), STITCH (3), SuperTarget (4) and the Druggable Genome database (5). These databases provide synoptic data on drugs and their primary or putative drug targets. Since the appearance of these drug/drug–target databases, other kinds of drug resources have emerged including PharmGKB (6), which specializes in pharmacogenetic and pharmacogenomic data, RxList (www.rxlist.com) and DailyMed (7), which provide electronic versions of the FDA's drug-product data sheets, ChEMBL (www. ebi.ac.uk/chembl) which provides data on drug-like compounds and BindingDB (8), which contains quantitative drug-binding constant data. The growing appetite for web-accessible drug data has also led PubChem (7),

KEGG (9), ChEBI (10) and ChemSpider (11) to add drugs and drug information to their usual offerings. All of these databases are outstanding resources, but as a general rule, most of them are quite 'lightly' annotated with only 10–15 data fields per drug entry.

In contrast to most other open-access drug databases, DrugBank (12) is a 'richly' annotated resource. The first version of DrugBank (released in 2006) contained nearly 90 data fields per drug entry, with detailed information on the nomenclature, ontology, chemistry, structure, function, action, pharmacology, pharmacokinetics, metabolism and pharmaceutical properties of both drugs and drug targets. Much of this data was acquired through primary literature sources, checked by experts, edited and entered manually. The richness, uniqueness and quality of the data in DrugBank has clearly hit a nerve with the research community. It is widely cited (more than 400 citations), integrated into many international databases (more than 20) and heavily used (more than 4 million page visits/year) by pharmacists, physicians, researchers, educators and the general public.

In an effort to keep up with the growing applications and far-ranging requests for this particular database, an updated version (DrugBank 2.0) was released in 2008 (13). Since then, the amount of easily accessible or predictable knowledge on drugs has grown considerably. So too has the number of requests, suggested improvements and calls for additional kinds of data to appear in DrugBank. Based on this user feedback we have spent the past 2 years enhancing both the quantity and quality of DrugBank's content. We have also added to or improved upon a number of DrugBank's querying and search functions. The net result is a 40% increase in the number of data fields for each drug entry, a considerable expansion (>50%) in the number of drug–protein and food–drug interactions, a massive increase in the information on drug metabolites, drug ADMET (absorption, distribution, metabolism, excretion, toxicology) data and the addition of hundreds of colorful, interactive, hand-drawn drug-action pathway diagrams. With these enhancements, we believe DrugBank has become a much more comprehensive and accessible drug information resource. It has also become significantly more useful for a wide range of 'omics' (i.e. pharmacogenomic, pharmacoproteomic, pharmacometabolomic, pharmacoeconomic) applications. A more detailed description of DrugBank 3.0 follows.

WHAT'S NEW IN DRUGBANK 3.0?

Details relating to DrugBank's overall design, general querying capabilities, curation protocols, structure depiction conventions, quality assurance and drug selection criteria have been described previously (12,13). These have largely remained the same between release 2.0 and 3.0. Here, we shall focus primarily on describing the changes and enhancements made to the database and to the annotation processes for release 3.0. More specifically, we will describe the: (i) growth and enhancements to DrugBank's existing content and quality; (ii) new data field additions; (iii) expanded database linkages and (iv) enhanced data querying and viewing capabilities.

GROWTH AND ENHANCEMENT OF EXISTING DATA

DrugBank has grown significantly in the past 5 years with perhaps the most significant changes happening between release 2.0 and 3.0. This progressive data content expansion is summarized in Table 1. As can be seen from this table, going from version 2.0 to 3.0, there has been a 40% increase in the number of data fields for each drug entry. Likewise there has been a 130% increase in the number of computed structure parameters, an 80% increase in the number of external database links, a 67% increase in the number of experimental drugs, a 46% increase in the number of food–drug interactions, a 42% increase in the total number of drug targets, a 20% increase in the number of possible DrugBank queries, a 13% increase in the number of FDA-approved drug targets, a 12% increase in the number of biotech and nutraceutical drugs and a 6% increase in the number of FDA-approved small molecule drugs.

In addition to significantly expanding the data content in DrugBank, a major effort has been directed at improving the quality of DrugBank's existing data. Hundreds of drug descriptions, mechanisms of action and pharmacological summaries have been either re-written or expanded. Likewise, hundreds of new drug and drug–target references were collected, checked and added. Similarly, extensive checks have been performed on all of DrugBank's small molecule structures to confirm that they exhibit the correct chirality and stereochemistry. In particular, we developed a custom structure-checking program that used direct structure comparison (via a Mol file) of each of DrugBank's structures against the corresponding structures in other databases (PubChem, ChEBI, ChemSpider, etc.). Any DrugBank structure that did not match with the corresponding structure in one or more of these external databases was flagged. A total of 340 structures were identified with potential structural errors or discrepancies. Each of these was assessed and/or corrected manually by a team of trained chemists. In many cases the DrugBank structure was correct and the external database structure was found to be in error, in other cases the DrugBank structure was determined to be in error and was subsequently corrected.

In addition to completing extensive data integrity checks for all of DrugBank's chemical data, the drug target information in DrugBank 3.0 has been significantly improved. Now most of DrugBank's approved-drug targets are prioritized by relevance, with each target being classified by its primary mode of action. One mode-of-action category lists targets known to confer the desired pharmacological effects, while the other lists targets with unknown or unintended pharmacological effects (many of which account for side effects). In addition to our implementation of an improved target classification scheme, DrugBank 3.0 now formally separates drug-action targets from drug transporters, drug

Table 1. Comparison between the coverage in DrugBank 1.0, 2.0 and DrugBank 3.0

Category	1.0	2.0	3.0
No. of data fields	88	108	148
No. of search types	8	12	16
No. of drug-action pathways	0	0	223
No. of drugs with metabolizing enzyme data	0	0	762
No. of drug metabolites	0	0	811
No. of drugs with drug transporter data	0	0	516
No. of SNP-associated drug effects	0	0	113
No. of drugs with patent/pricing/manufacturer data	0	0	1208
No. of food–drug interactions	0	714	1039
No. of drug–drug interactions	0	13 242	13 795
No. of ADMET parameters (Caco-2, LogS)	0	276	890
No. of QSAR parameters per drug	5	6	14
No. of FDA-approved small molecule drugs	841	1344	1424
No. of biotech drugs	113	123	132
No. of nutraceutical drugs	61	69	82
No. of withdrawn drugs	0	57	68
No. of illicit drugs	0	188	189
No. of experimental drugs	2894	3116	5210
Total No. of experimental and FDA small molecule drugs	3796	4774	6684
Total No. of experimental and FDA drugs	3909	4897	6816
No. of names/brands/synonyms	18 304	28 447	37 171
No. of approved-drug drug targets (unique)	524	1565	1768
No. of all drug targets (unique)	2133	3037	4326
No. of approved-drug enzymes/carriers (unique)	0	0	164
No. of all drug enzymes/carriers (unique)	0	0	169
No. of external database links	12	18	31

carriers and pro-drug conversion enzymes. Note that in DrugBank, carriers are considered separate from transporters as carriers move drugs around the body, while transporters move drugs into and around the cell. This kind of target separation should make drug-target studies somewhat easier and substantially more informative.

While many visible 'front-end' enhancements have been implemented, DrugBank's back-end has also been significantly enhanced. In particular, all of DrugBank's data has been converted to an easily parsed XML format. This should make data downloads and the development of data extraction routines much simpler and far faster for programmers and database developers.

NEW DATA FIELD ADDITIONS

Each new release of DrugBank has been characterized by a significant increase in the number of new data fields compared to the previous release. DrugBank 3.0 is no exception. Going from version 2.0 to 3.0, there has been a substantial increase in the number of data fields (going from 108 to 148). Many of these data field additions were the results of specific requests by DrugBank users or arose through consultation with members of the pharmaceutical research community. DrugBank 3.0 now includes drug pathway diagrams, drug transporter information, drug carrier information, drug metabolite data, drug metabolizing enzyme data, QSAR data, chemical classification data, SNP-associated drug effects (available through the GenoBrowse link) and drug patent/pricing/ manufacturer data. Table 2 provides a more complete

listing of the new data fields appearing in DrugBank 3.0. The five areas where most of the new data has been added relate to: (i) pharmacometabolomics; (ii) pharmacoproteomics; (iii) pharmacogenomics; (iv) pharmacoeconomics and (v) computed structure features. These additions are detailed below:

Pharmacometabolomics

As the field of metabolomics has grown, so too has the interest in understanding drug metabolism and in characterizing drug metabolites. Indeed, it is now recognized that drug metabolites play an important role in understanding adverse drug effects, in determining therapeutic indices and in leading to secondary or off-target therapeutic effects. In an effort to make this kind of metabolic or metabolomic information readily available we have manually compiled detailed phase I/II metabolic fate data for over 760 FDA-approved drugs. The data, which was compiled from hundreds of journal articles, includes links to pathways, chemical structures, HMDB (14) entries, reaction parameters such as reaction type (inducer, inhibitor and substrate), K_m, and V_{max}, and reaction class information for nearly 720 drugs. In addition to these annotations and molecular descriptors for drug metabolism and drug metabolites, DrugBank 3.0 also provides classical ADMET information including drug distribution, clearance and route of elimination data. These new 'metabolic' data fields, when combined with the structural and biological information already stored in DrugBank, should offer an entirely new way of exploiting the data in DrugBank.

Table 2. New data fields in DrugBank 3.0

Chemical kingdom	ChEMBL link
Chemical class	Drug pathway
Chemical substructures	Drug pathway SMPDB ID
Drug manufacturers	Target actions (antagonist, agonist)
Drug packagers	Target priority
Drug prices	Target pharmacological effect (known/unknown/none)
Original patent date	Enzyme actions (inhibitor, inducer, substrate)
Original patent number	Drug metabolite structure
Last patent expiry date	Drug metabolite name
Last patent number	Drug metabolite HMDB ID
MMCD link	Drug metabolite reaction type (e.g. oxidation)
ChemSpider ID and link	Reaction K_m value
NDC ID and link	Reaction V_{max} value
DailyMed link	Metabolizing enzyme references
Drugs.com link	Metabolizing enzyme priority
OMIM link	Drug transporter name
CDPD link	Drug transporter actions (substrate, inhibitor, inducer)
TTD link	Route of elimination
STITCH link	Volume of distribution
BindingDB link	Clearance

Pharmacoproteomics

It is often said that a picture is worth a thousand words. In order to simplify DrugBank's vast collection proteomic data, DrugBank 3.0 now includes nearly 230 richly illustrated drug-action pathways. These pathways have been designed to display the action of drugs on protein targets or protein receptors. Using the visualization framework developed for SMPDB (15), each DrugBank pathway is 'image-mapped', with every drug structure being hyperlinked to the detailed descriptions contained in DrugBank or HMDB and every protein or enzyme complex being hyperlinked to the detailed descriptions provided by UniProt (16). DrugBank's drug-action pathways are carefully hand-drawn and frequently include information on the relevant organs, organelles, subcellular compartments, protein targets, protein locations and drug structures that describe the pharmacology or mode of action for that drug. All of DrugBank's pathways images may be progressively expanded by clicking on the Zoom button located at the top and bottom of the image or the magnifying-glass icons in the Highlight/Analyzer box on the right of the image. At the top of each image is a pathway synopsis while at the bottom of each image is a list of relevant references. Over the coming 2–3 years it is expected that another 1000–1200 drug pathways will be added to DrugBank's pathway inventory.

Pharmacogenomics

The relationship between drugs, genes and genetic variants (SNPs) is central to the whole field of pharmacogenomics and personalized medicine. In an effort to address these rapidly growing needs, DrugBank 3.0 now contains a significant amount of new pharmacogenomic information, including data on 26 292 coding (exon) SNPs and 73 328 non-coding (intron) SNPs derived from known drug targets. It also has data on 1188 coding SNPs and 8931 non-coding SNPs from known drug metabolizing enzymes. SNP information can be accessed by clicking on the 'Show SNPs' hyperlink listed beside either the metabolizing enzymes or the drug target SNP field. These SNP summary tables include: (i) the reference SNP ID, with a hyperlink to dbSNP (17); (ii) the allele variants; (iii) the validation status; (iv) the chromosome location and reference base position; (v) the functional class (synonymous, non-synonymous, untranslated, intron, exon); (vi) mRNA and protein accession links (if applicable); (vii) the reading frame (if applicable); (viii) the amino acid change (if existent); (ix) the allele frequency as measured in African, European and Asian populations (if available) and (x) the sequence of the gene fragment with the SNP highlighted in a red box. The purpose of these SNP tables is to allow DrugBank users to go directly from a drug of interest to a list of potential SNPs that may contribute to the reaction or response seen in a given patient or in a given population.

In addition to this drug target SNP data, Drugbank 3.0 now includes two tables that provide much more explicit information on the relationship between drug responses/ reactions and gene variant or SNP data. These tables, which are accessible from the GenoBrowse submenu located on DrugBank's Browse menu bar, are called SNP-FX (short for SNP-associated effects) and SNP-ADR (short for SNP-associated adverse drug reactions). SNP-FX contains data on the drug, the interacting protein(s), the 'causal' SNPs or genetic variants for that gene/protein, the therapeutic response or effects caused by the SNP-drug interaction (improved or diminished response, changed dosing requirements, etc.) and the associated references describing these effects in more detail. SNP-ADR follows a similar format to SNP-FX but the clinical responses are restricted only to adverse drug reactions (ADR). SNP-FX contains literature-derived data on the therapeutic effects or therapeutic responses for more than 60 drug-polymorphism combinations, while SNP-ADR contains data on adverse reactions compiled from more than 50 drug-polymorphsim pairings. All of the data in these tables is hyperlinked to drug entries from DrugBank, protein data from UniProt, SNP data from dbSNP and bibliographic data from PubMed.

Pharmacoeconomics

Perhaps the most important 'omics' discipline in the pharmaceutical world is econ-'omics'. Indeed, the selection of disease targets, the money spent on research and the money recovered from sales are largely determined by economic factors. To facilitate research into these issues DrugBank 3.0 has added data on drug patent dates (from Canada and the United States), drug manufacturers, drug packagers, drug prices (from different jurisdictions) and drug sales (where available). Coupled with other data already in DrugBank (such as ATC codes, indications, side effects, structures, chemical classes) this information should enable more detailed studies on the relationship between drug prices and patent dates, the connection

between drug prices and drug sales, the relationship between drug sales and disease targets, the link between a drug's price and the drug's side effects as well as the relationship between drug manufacturers and disease target choices. It should also enable research into historical trends in drug target or disease target choices as well as long-term trends in the use or exploitation of certain drug motifs or structure classes (e.g. statins, tricyclic drugs).

Computed structure data

Computed structure parameters or descriptors are frequently used in quantitative structure activity relationship (QSAR) studies to facilitate rational drug design, drug screening and medicinal chemistry. These computed structure parameters may also be used to rationalize drug activities, tissue localization, adverse reactions and drug metabolism. While earlier versions of DrugBank had provided a nominal number of computed structure descriptors (molecular weight, pKa, LogP, LogS), these were often not sufficient for detailed feature analyses or comprehensive *in silico* comparisons. DrugBank 3.0 now provides an additional set of nine computed property descriptors including (i) the number of H-bond acceptors; (ii) the number of H-bond donors; (iii) the number of freely rotating bonds; (iv) the index of refraction; (v) the predicted boiling/melting point; (vi) the polar surface area; (vii) the molar refractivity; (viii) the polarizability and (ix) the molecular density. These properties are displayed in a summary table with links to the SDF file containing these values. While more computed structure descriptors are certainly available, these represent the most frequently used descriptors and should allow DrugBank users to perform much more detailed computed structure queries, analyses and comparisons.

EXPANDED DATABASE LINKAGES

Because DrugBank was designed to cover a broad spectrum of scientific disciplines it has always been extensively linked to many external databases. For instance, version 2.0 of DrugBank contained up to 18 database hyperlinks in every DrugCard entry, including links to KEGG (9), PubChem (7), ChEBI (10), PharmGKB (6), PDB (18), GenBank (19), DIN, RxList, PDRhealth, Wikipedia, ATC, UniProt (16), Pfam (20), dbSNP (17), GeneCards (21), GenAtlas (22), HGNC and PubMed. DrugBank 3.0 now contains an average of 31 hyperlinks per DrugCard. These new links include numerous compound-specific, spectral, pathway and disease databases such as ChemSpider (11), HMDB (14), MMCD (23), SMPDB (15) and OMIM (24). We have also added new links to several dedicated drug and pharmaceutical databases [DailyMed (7), Drugs.com, the National Drug Code identifier database and the Canadian Drug Product Database] as well as a number of drug target databases, such as the Therapeutic Target Database (TTD), STITCH (4), BindingDB (8) and ChEMBL. These DrugCard hyperlinks are also complemented with a comprehensive

list of links in the 'About' section of DrugBank. In addition to these external database links, DrugBank has been reciprocally linked to several major resources including Wikipedia, UniProt (16), BioMOBY (25), PubChem (7), KEGG (9), PharmGKB (6), Drugs.com and ChemSpider (11).

ENHANCED QUERYING AND VIEWING CAPABILITIES

One of the key strengths for DrugBank has been its support for a wide range of querying and visualization tools. These include 2D and 3D structure viewers, flexible text querying systems, structure searching/matching, sequence searching tools, data extraction tools and easy-to-use browsers. For DrugBank 3.0 we have made a number of improvements to the existing query tools but are also introducing four new browsing or search tools. These include PathBrowse, GenoBrowse, ClassBrowse, ReceptorBrowse and the Interax Interaction Search (Figure 1). We believe all five should make the viewing and retrieval of information in DrugBank much easier.

PathBrowse was developed to facilitate the viewing and searching of DrugBank's drug-action pathways. Each hyperlinked, interactive pathway explains the mode of action of drugs at a molecular, cellular and/or physiological level. PathBrowse allows users to search for drugs by DrugBank ID, name or synonyms. It also supports the search for drug targets, metabolizing enzymes, carriers and transporters either by their name, UniProt ID or gene identifier. The results are displayed as a highlighted list of hits. Once a pathway is selected, users can interactively explore the pathway image, with compound or protein hits highlighted in the pathway image. This tight integration between DrugBank and SMPDB should allow researchers to visualize the 'big picture' with respect to drugs and how they act or how they are processed in the body. The two other browsing functions (ClassBrowse and GenoBrowse) are somewhat simpler in design and functionality than PathBrowse. ClassBrowse allows users to search through or sort drugs by their chemical class or chemical taxonomy while GenoBrowse (which has already been described) allows users to browse through or explore SNP-induced drug effects or drug reactions. ReceptorBrowse allows users to search or sort through the protein targets, enzymes, carriers and transporters (along with their function and target species information) that are associated with each drug in DrugBank.

DrugBank contains one of the most complete, freely available sources of drug–drug and food–drug interaction data on the Internet today. Although this information has been made available in each DrugCard from version 2.0 onwards, the data has not been easily searchable. The 'Interax' Interaction Search was developed to allow facile searching of drug and food interactions. Unlike existing interaction search tools, Interax takes the process one step further by including transporter, target and enzyme information in the search results. Several different search types are supported by Interax. For instance,
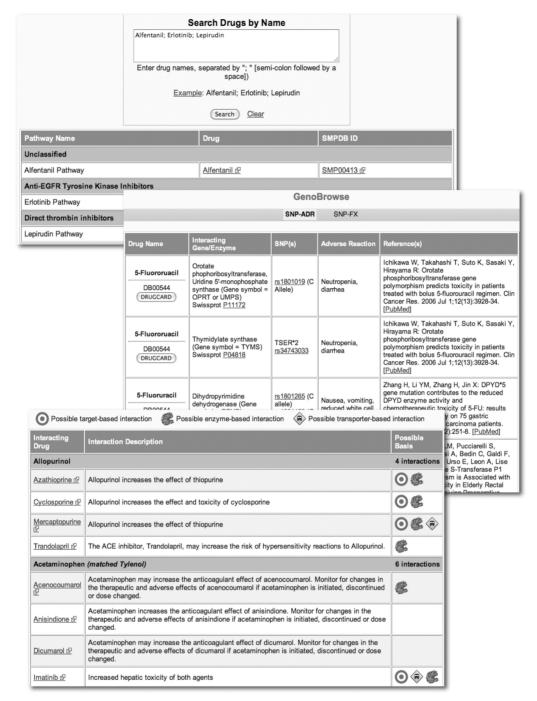
Figure 1. A screenshot montage of some of DrugBank 3.0's new browser views (GenoBrowse, Interax, ClassBrowse, PathBrowse).

standard drug–drug or food–drug interaction searches can be performed, whereby a user inputs a list of drugs, presses the 'submit' button and a list of drug and food interactions are produced. Users can also input two lists of drugs and Interax will identify any interactions between the lists. Additionally, any interactions that may be target, enzyme, carrier or transporter related (e.g. two drugs bind the same target) will be flagged with symbols representing a target interaction, enzyme interaction, carrier interaction or transporter interaction. This comprehensive search functionality provides a unique method of searching and exploring drug–drug interactions and should be of interest to pharmacists, pharmaceutical researchers, and the general public.

CONCLUSION

DrugBank 3.0 contains a significant number of enhancements over its predecessor (DrugBank 2.0). As highlighted throughout this article, numerous improvements have

been made in the quantity, quality, depth and organization of the information provided. These include the addition of new drugs, new targets, new data fields, new links and new tools. DrugBank 3.0 now contains illustrated drug-action pathways, drug transporter data, drug metabolite data, pharmacogenomic data, adverse drug response data, ADMET data, pharmacokinetic data, extensive computed property data and chemical classification data. DrugBank 3.0 also offers expanded database links, improved search tools for drug–drug and food–drug interaction, new tools for searching and viewing drug pathways and hundreds of new drug entries with detailed patent, pricing and manufacturer data. These additions have been complemented by enhancements to the quality and quantity of existing data, particularly with regard to drug target, drug description and drug action data. With these enhancements DrugBank 3.0 should be much more useful for a wider range of 'omics' applications. It is hoped that with more user feedback, DrugBank will continue to develop to fit the needs of its users and provide an increasingly useful, information-rich drug resource.

REFERENCES

1. Zhu,F., Han,G., Kumar,P., Liu,X., Ma,X., Wei,X., Huang,L., Guo,Y., Han,L., Zheng,C. *et al.* (2010) Update of TTD: therapeutic target database. *Nucleic Acids Res.*, **38(Database issue)**, D787–D791.
2. Gao,Z., Li,H., Zhang,H., Liu,X., Kang,L., Luo,X., Zhu,W., Chen,K., Wang,X. and Jiang,H. (2008) PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics*, **9**, 104.
3. Kuhn,M., Szklarczyk,D., Franceschini,A., Campillos,M., von Mering,C., Jensen,L.J., Beyer,A. and Bork,P. (2010) STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Res.*, **38(Database issue)**, D552–D556.
4. Gunther,S., Kun,M., Dunkel,M., Camillos,M., Senger,C., Petsalaki,E., Ahmed,J., Urdiales,E.G., Gewiess,A., Jensen,L.J. *et al.* (2008) SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.*, **36(Database issue)**, D919–D922.
5. Russ,A.P. and Lampel,S. (2005) The druggable genome: an update. *Drug Discov. Today*, **10**, 1607–1610.
6. Hernandez-Boussard,T., Whirl-Carrillo,M., Hebert,J.M., Gong,L., Owen,R., Gong,M., Gor,W., Liu,F., Truong,C., Whaley,R. *et al.* (2008) The pharmacogenetics and pharmacogenomics knowledge base: accentuating the knowledge. *Nucleic Acids Res.*, **36(Database issue)**, D913–D918.
7. Wang,Y., Xiao,J., Suzek,T.O., Zhang,J., Wang,J. and Bryant,S.H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.*, **37(Database issue)**, W623–W633.
8. Liu,T., Lin,Y., Wen,X., Jorissen,R.N. and Bilson,M.K. (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.*, **35(Database issue)**, D198–D201.
9. Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M. and Hirakawa,M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38(Database issue)**, D355–D360.
10. Brooksbank,C., Cameron,G. and Thornton,J. (2005) The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res.*, **33(Database issue)**, D46–D53.
11. Williams,A.J. (2008) Public chemical compound databases. *Curr. Opin. Drug Discov. Devel.*, **11**, 393–404.
12. Wishart,D.S., Knox,C., Guo,A.C., Shrivastava,S., Hassanali,M., Stothard,P., Chang,Z. and Woolsey,J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34(Database issue)**, D668–D672.
13. Wishart,D.S., Knox,C., Guo,A.C., Cheng,D., Shrivastava,S., Tzur,D., Gautam,B. and Hassanali,M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36(Database issue)**, D901–D906.
14. Wishart,D.S., Knox,C., Guo,A.C., Eisner,R., Young,N., Gautam,B., Hau,D.D., Psychogios,N., Bouatra,S., Mandal,R. *et al.* (2009) HMDB: a knowledgebase for the human meabolome. *Nucleic Acids Res.*, **37(Database issue)**, D603–D610.
15. Frolkis,A., Knox,C., Lim,E., Jewison,T., Law,V., Hau,D.D., Liu,P., Gautam,B., Ly,S., Guo,A.C. *et al.* (2010) SMPDB: the small molecule pathway database. *Nucleic Acids Res.*, **38(Database issue)**, D480–D487.
16. UniProt Consortium. (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37(Database issue)**, D169–D174.
17. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
18. Kouranov,A., Xie,L., de la Cruz,J., Chen,L., Westbrook,J., Bourne,P.E. and Berman,H.M. (2006) The RCSB PDB information portal for structural genomics. *Nucleic Acids Res.*, **34(Database issue)**, D302–D305.
19. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2010) GenBank. *Nucleic Acids Res.*, **38(Database issue)**, D46–D51.
20. Finn,R.D., Mistry,J., Tate,J., Coggill,P., Heger,A., Polliington,J.E., Gavin,O.L., Gunasekaran,P., Ceric,G., Froslund,K. *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38(Database issue)**, D211–D222.
21. Safran,M., Dalah,I., Alexander,J., Rosen,N., Iny Stein,T., Shmoish,M., Nativ,N., Bahir,I., Doniger,T., Krug,H. *et al.* (2010) GeneCards Version 3: the human gene integrator. *Database*, doi: 10.1093/baq020; [7 August 2010, epub ahead of print].
22. Roux-Rouquie,M., Chavet,M.L., Munnich,A. and Frezal,J. (1999) Human genes involvedin chromatin remodeling in transcription initiation and associated diseases: an overview using the GENATLAS database. *Mol. Genet. Metab.*, **62**, 261–277.
23. Markley,J.L., Anderson,M.E., Cui,Q., Eghbalnia,H.R., Lewis,I.A., Hegeman,A.D., Li,J., Schulte,C.F., Sussman,M.R., Westler,W.M. *et al.* (2007) New bioinformatics resources for metabolomics. *Pac. Symp. Biocomput.*, **12**, 157–168.
24. Hamosh,A., Scott,A.F., Amberger,J.S., Bocchini,C.A. and McKusick,V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33(Database issue)**, D514–D517.
25. Kawas,E., Senger,M. and Wilkinson,M.D. (2006) BioMoby extensions to the Taverna workflow management and enactment software. *BMC Bioinformatics*, **7**, 523.